VALUE IN HEALTH 18 (2015) 856–864

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Calculating Preference Weights for the Labor and Delivery Index: A Discrete Choice Experiment on Women's Birth Experiences

Fania R. Gärtner, PhD[1,], Esther W. de Bekker-Grob, PhD[2], Anne M. Stiggelbout, PhD[1], Marlies E. Rijnders, PhD[3], Liv M. Freeman, MD[4], Johanna M. Middeldorp, MD, PhD[4], Kitty W.M. Bloemenkamp, MD, PhD[4], Esteriek de Miranda, PhD[5], M. Elske van den Akker-van Marle, PhD[1]*

[1]*Department of Medical Decision Making, Leiden University Medical Center, Leiden, The Netherlands;* [2]*Department of Public Health, Erasmus MC University Medical Center, Rotterdam, The Netherlands;* [3]*TNO Child Health, Leiden, The Netherlands;* [4]*Department of Obstetrics, Leiden University Medical Center, Leiden, The Netherlands;* [5]*Department of Gynecology and Obstetrics and Gynecology, Academic Medical Center, Amsterdam, The Netherlands*

ABSTRACT

Objective: The aim of this study was to calculate preference weights for the Labor and Delivery Index (LADY-X) to make it suitable as a utility measure for perinatal care studies. **Methods:** In an online discrete choice experiment, 18 pairs of hypothetical scenarios were presented to respondents, from which they had to choose a preferred option. The scenarios describe the birth experience in terms of the seven LADY-X attributes. A D-efficient discrete choice experiment design with priors based on a small sample (N = 110) was applied. Two samples were gathered, women who had recently given birth and subjects from the general population. Both samples were analyzed separately using a panel mixed logit (MMNL) model. Using the panel mixed multinomial logit (MMNL) model results and accounting for preference heterogeneity, we calculated the average preference weights for LADY-X attribute levels. These were transformed to represent a utility score between 0 and 1, with 0 representing the worst and 1 representing the best birth experience. **Results:** In total, 1097 women who had recently given birth and 367 subjects from the general population participated. Greater value was placed on differences between bottom and middle attribute levels than on differences between middle and top levels. The attributes that resulted in larger utility increases than the other attributes were "feeling of safety" in the sample of women who had recently given birth and "feeling of safety" and "availability of professionals" in the general population sample. **Conclusions:** By using the derived preference weights, LADY-X has the potential to be used as a utility measure for perinatal (cost-) effectiveness studies.

Keywords: birth experience, preferences, tariff, utility.

Introduction

To improve birth outcomes, new medical interventions are constantly evaluated against routine care. Several birth-specific outcomes or domains are typically used in perinatal studies, an approach that can lead to contradictory results and that often does not reflect what is important to women [1,2]. Examples of outcomes used are pain [3,4], labor duration [5], or anxiety [6,7]. A birth-specific utility measure reflecting the course of labor and birth would be more informative, as Petrou et al. [8] called for in 2003. Unfortunately, the use of generic quality-of-life utility measures, such as the EuroQol five-dimensional questionnaire [9], is regarded as less appropriate for perinatal care studies [8]. Because of the relatively short duration of labor and birth, differences between interventions affecting only the course of labor and birth, and not the long-term health of the child, will not be reflected in quality-of-life measures. Furthermore, items of existing health-related quality-of-life instruments are considered too crude to capture the aspects of health that are assumed to be affected in the perinatal phase.

To satisfy the need for a utility outcome measure for the intrapartum care setting, we developed a new birth-specific utility measure. In a mixed-methods study, we developed a set of seven birth-specific domains to include in such a measure [10]. On the basis of these domains, we formulated items and response categories that comprise the new questionnaire, the Labor and Delivery Index (LADY-X). LADY-X is a subjective evaluation of labor and delivery by domains that reflect the key aspects of a mother's overall experience of labor and birth. To make LADY-X suitable as a utility measure, a scoring algorithm is

Conflict of interest: The authors have indicated that they have no conflicts of interest with regard to the content of this article.

* *Address correspondence to:* Fania R. Gärtner, Department of Medical Decision Making, Leiden University Medical Center, Postbus 9600, 2300RC Leiden, The Netherlands.

E-mail: f.r.gartner@lumc.nl.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2015.07.005

necessary to calculate preference weights for each birth situation as classified by LADY-X. Therefore, the aim of this study was to estimate these preference weights.

For this purpose, we performed a discrete choice experiment (DCE). DCEs are based on the assumption that the object of interest (in the health care context, these are typically health care services, interventions, policies, or outcomes) can be described by a set of domains (attributes) to which levels are assigned and on which the individual's valuation depends. DCEs are intended to indirectly elicit patient preferences for alternative combinations of attribute levels by presenting a number of choice sets that each consist of two or more different combinations of attribute levels. Respondents compare the different alternatives of one set and choose the one with the highest latent utility value [11–13]. DCEs are used in health care research for various purposes, such as valuating patient experiences of health care interventions, valuating health outcomes, or assessing trade-offs between health outcomes and patient experiences [14,15], including in intrapartum research [16–20]. In the past decade, DCEs have also been applied to estimate preference weights (utilities) for classification systems [15].

Whether the patients' or the public's values are considered relevant for the estimation of preference weights is subject of ongoing debate [21–27]. Because no consensus yet exists, we estimated two sets of preference weights for LADY-X, one based on a sample of women who recently gave birth (henceforth referred to as the "women sample") and one based on a sample of the general population (men and women 18 years and older).

Methods

The Labor and Delivery Index

LADY-X measures the labor and delivery experience of the woman who gave birth. It consists of seven items: 1) availability of competent professionals, 2) the information provided, 3) professionals' responses to needs, 4) professionals' emotional support, 5) feelings of safety, 6) concerns about the child's condition, and 7) duration until first contact with child. Each item has three response categories, which vary by item; however, for all items, we presumed an ordering of the levels ranging from "very well," "adequately," to "inadequately." The levels of the last item "duration until first contact" vary slightly from this categorization. Based on these labels, the distance between the bottom category and the middle category is considered more meaningful than the distance between the middle category and the upper category for items from one to six. For an overview of the seven items (termed attributes in the DCE context) and their levels, see Table 1.

The seven items were derived from a mixed-methods study that included the views of pregnant women, women who had recently given birth, and professionals analyzing which aspects of labor and delivery are most important for a mother's overall experience of labor and birth [10]. The items were further evaluated in eight verbal probe interviews with women who had given birth in the past year. On the basis of these interviews, we concluded that the seven domains of LADY-X are clear, distinctive, relevant, complete, and applicable to all types of birth (place of birth, cesarean section, occurrence of complications, etc.). Thus, we concluded that LADY-X has good content validity. Parallel to this study intended to estimate preference weights for LADY-X, a clinometric evaluation of LADY-X was performed, which exhibits good reliability and construct validity. Results of the validation study will be presented elsewhere. An English version of LADY-X based on a forward-backward translation process is presented in the Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.07.005. The English version is yet to be evaluated in terms of validity and reliability.

Discrete Choice Design

The assumption in the DCE is that the object (e.g., a health care intervention or program) that is being valued is defined by a number of characteristics (attributes) and levels that are assigned to these attributes, as is the case for classification systems. In the present study, the object of interest was the birth experience of a mother, the seven items of LADY-X were the attributes that define the birth experience, and the three response categories of each item (e.g., "very well," "adequately," and "inadequately") formed the attribute levels (see Table 1). The relative importance of these seven attributes was assessed by presenting respondents with a series of choice sets consisting of two hypothetical scenarios with varying combinations of attribute levels (see Figure 1 for an example of a choice set). For each choice set, respondents were asked to indicate their preferred birth experience scenario [28].

In this study, we had seven attributes with three levels each, and thus $3^7 = 2137$ possible birth experience scenarios. Because it is not feasible to present a single individual with all these scenarios, a sample of 36 scenarios out of all possible scenarios was selected [29]. To further reduce the burden on respondents [30], we divided the 36 scenarios into two questionnaires using a blocked design with near attribute level balance [28].

For the selection of the scenarios and choice sets, we applied a D-efficient DCE design with priors [29] using Ngene software (version 1.1.1, http://www.choice-metrics.com/). That is, our DCE design consisted of two phases. In the first phase, we created a DCE design using zero priors (i.e., input information). This design was sent to a relatively small sample (N = 110). In the second phase, estimated parameters from the first phase were used as priors to create a more efficient DCE design. Such an efficient DCE design with priors is considered to improve the reliability of the results and leads to smaller standard errors [31–33]. Using this

Table 1 – Attribute description and levels.

	Attributes		Level		
	Name	Abbreviation	1	2	3
1	Availability of competent health care professionals	Availability	At all times	Most of the time	Rarely
2	Information provided by health care professionals	Information	Very well	Adequately	Inadequately
3	Health care professionals' responses to needs	Needs	Very	Reasonably	Not at all
4	Emotional support by health care professionals	Emotional support	Very well	Adequately	Inadequately
5	The mother's feelings of safety	Safety	Very	Reasonably	Not enough
6	The mother's concerns about the child's condition	Concerns	No	Some	Many
7	Experienced duration until first contact with the child	First contact	Not long	Quite long	Very long

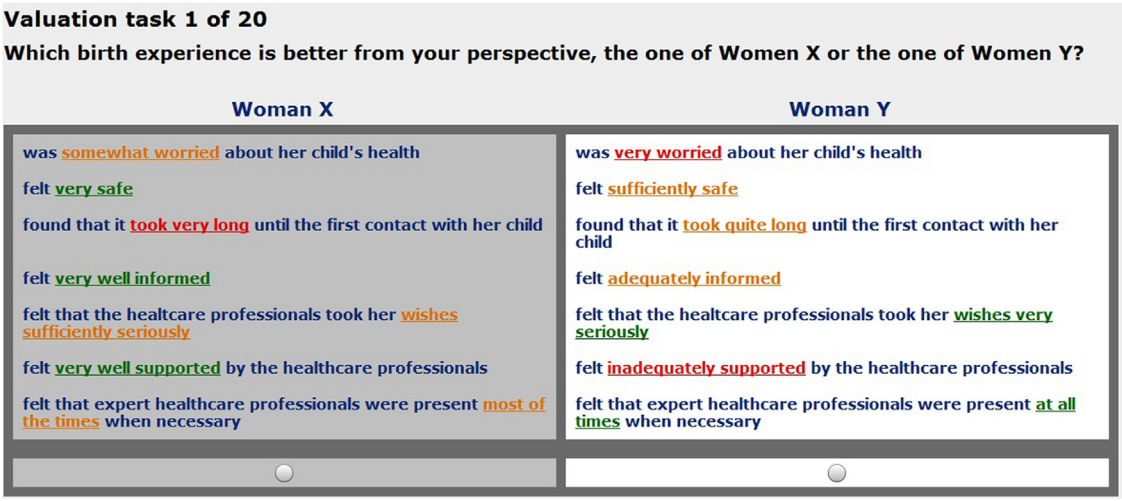

Fig. 1 – Example of the choice task. (Color version of figure is available online.)

design, consisting of two questionnaire versions with 18 choice sets each containing two birth experience alternatives each, we were able to estimate at least all main effects.

Questionnaire Design

The DCE was presented in an online questionnaire. Each questionnaire began with detailed information on the attributes and attribute levels and included a clearly explained example of a choice set containing two birth experience scenarios. The main part of each questionnaire comprised the 18 choice sets. We labeled the two scenarios as the experience of Woman X and the experience of Woman Y, and we asked the subject to "select the woman who from your perspective had the better birth experience, Woman X or Woman Y." Based on pretesting (see the next subheading), the attribute levels were colored to reduce the complexity of the choice sets: green for the most positive levels, orange for the intermediate levels, and red for the worst levels.

To obtain insights into the reliability and rationality of respondents' choices, two choice sets with a dominant birth experience scenario were added to the choice tasks; that is, one scenario was better with respect to all levels and therefore should be chosen by all respondents who comprehend the experiment and answer the questions seriously. Those respondents who did not answer these two choices correctly were defined as inconsistent and dropped from the analyses [13].

The order in which the attributes were presented in the alternatives was randomly determined for each respondent, to avoid order effects, but for each respondent, the order was consistent over all 20 choice tasks (18 plus 2 consistency choice sets).

Pretesting of the questionnaire

Before the data collection, the questionnaire was pretested during 10 verbal probe interviews [34], five with women who had recently given birth and five with representatives of the general population. Our aim in the pretest was to improve the layout and phrasing of the DCE and obtain an indication of the feasibility of the questionnaire.

In the 1.5-hour interviews, the questionnaire was presented to the respondents and we asked them to probe any thoughts, difficulties, and ideas for improvement and to clarify the reasoning behind the choices they made in the choice tasks. In addition, the interviewer interrogated about the clarity of the instructions, attributes, and levels and the layout of the choice tasks including use of colors and a three- versus two-column layout. Moreover, questions were asked regarding cognitive and motivational aspects, such as the motivation to complete the task and specific, less-prevalent, and thus likely rather unrealistic combinations of attribute levels. In total, we tested eight of these combinations explicitly by including them in the pretest questionnaire, and after the respondent had completed the specific choice task probe, the respondent was asked to report any difficulties associated with or specific thoughts about it. These eight combinations were selected on the basis of a sample of data we collected for the validation study of LADY-X at the time of the pretest (N = 169) and for which level 1 of one attribute and level 3 of another attribute never occurred together for one respondent.

On the basis of the pretest, we concluded that respondents were able to understand the choice task and make choices on the basis of their personal opinions, including subjects who had never or a long time ago had given birth or attended a birth. Overall, respondents did not become confused by any unlikely combinations of attribute levels, and none of these combinations hindered the respondents in making their choices. This confirms that the use of an unrestricted DCE design was justified because using a restricted design might have a negative impact on the efficiency level of the DCE design. According to the respondents, no important attributes were missing in the DCE. Most of the subjects felt confident in their choices, and they imagined both home and hospital birth settings during the choice task. In the pretest, respondents were able to address 20 choice tasks, whereas addressing 26 was considered excessive and reduced the motivation and concentration of respondents. As a result of the pretest, we shortened the instruction text and adapted the layout. We retained the color coding of the attribute levels (green, orange, and red).

Participants and Data Collection

This study included two samples, one of women who had recently given birth and one of the general population. To be included in the women sample, women had to have given birth between July 2012 and January 2014 and be conversant in Dutch. We imposed no restriction on any obstetric characteristic. We recruited women through advertisements in regional newspapers and through communication channels of one Dutch academic medical center, that is, requests on the intranet, the official Web site, and posts on its Twitter and Facebook accounts. These posts were shared many times, including by primary-care midwives,

and thus reached women throughout the country. Furthermore, 632 women who had participated in earlier studies conducted by the research team and had professed an interest in follow-up studies were invited to complete the questionnaire. Unique links to the online questionnaire were e-mailed to the respondents, and a reminder was sent 1 week later to those who had not replied. The Medical Ethics Committee of the Leiden University Medical Center approved this study.

To collect the data for the general population sample, an online survey company distributed unique links to the questionnaire to a representative sample of adults (18 years or older) in The Netherlands. To guarantee representativeness, we applied quotas to control for the ethnic background, sex, and age distribution of the sample.

We administered the (on average 20-minute) online questionnaires in December 2013 (DCE design with zero priors; first phase) and in January 2014 (DCE design with prior information; second phase). For the women sample, it also included questions on obstetrical background (number of pregnancies, miscarriages, and parity), facts concerning the most recent birth (place of birth, method of birth, single vs. multiple pregnancy, use of pain relief during labor, complications during labor, maternal and neonatal morbidity, and LADY-X).

Analyses

Data sets from the first-phase DCE (without prior information) and the second-phase DCE (with prior information) were combined—separately for the women and general population samples—and analyzed using a panel mixed logit (MMNL) model to account for preference heterogeneity and correlation between the 18 choice tasks completed by each individual [28]. A normal distribution was applied to all attributes. Based on the Akaike information criterion, this MMNL model had a better fit than a more restrictive multinomial logit model or a latent class model. After testing for left-right bias, the following utility model was estimated under the assumption that all parameters were normally distributed:

$$\begin{aligned} V = {} & \beta_1 \text{Availability(at\_all\_times)} + \beta_2 \text{Availability(most\_of\_the\_time)} + \\ & \beta_3 \text{Information(very\_well)} + \beta_4 \text{Information(adequately)} + \\ & \beta_5 \text{Needs(very\_well)} + \beta_6 \text{Needs(reasonably)} + \\ & \beta_7 \text{Support(very\_well)} + \beta_8 \text{Support(adequately)} + \\ & \beta_9 \text{Safety(very\_well)} + \beta_{10} \text{Safety(reasonably)} + \\ & \beta_{11} \text{Concerns(no)} + \beta_{12} \text{Concerns(some)} + \\ & \beta_{13} \text{First\_contact(not\_long)} + \beta_{14} \text{First\_contact(quite\_long)} \end{aligned} \quad (1)$$

V is the observable relative utility of a birth experience scenario that is composed of the individual β coefficients of the model. β_1 to β_{14} are coefficients of the attributes indicating the relative weight individuals place on a certain attribute level. The statistical significance of a coefficient ($P \leq 0.05$) indicates that individuals differentiated between one attribute level and another in making stated choices. The middle level of each attribute served as the reference level. Therefore, we expected all upper attribute levels to be positive (i.e., a positive sign) and the lowest attribute levels to be negative (i.e., a negative sign). Analyses were performed in Nlogit 5.0 (Econometric Software Inc., Plainview, NY, USA).

Estimation of LADY-X Preference Weights

For the final calculation of LADY-X preference weights, parameter values (coefficients) of the panel MMNL were used. We calculated preference weights separately for both samples. We used the method described by Hoefman et al. [35]. In brief, we first reconstructed the random parameter distribution from our sample because the parameters of the panel MMNL represent population-level estimates (unconditional distribution) [28]. Using bootstrap sampling, we randomly assigned this unconditional distribution over a hypothetical sample of 10,000 individuals. The resulting individual-specific parameter estimates were averaged and rescaled to represent LADY-X preference weights. This was done by 1) adding up the mean parameter values (coefficients) of the best LADY-X birth experience (i.e., the best response category for each of the seven LADY-X items) and 2) dividing all estimates by this total score [35]. Next, these relative scores were transformed to convert preference weights (from the MMNL model) to a more intuitive scale [36]. We assigned 0 to the worst birth experience scenario (i.e., the birth experience scenario, with all LADY-X items having the lowest possible score) and 1 to the top birth experience scenario (i.e., the birth experience scenario, with all LADY-X items having the highest possible score) [37]. An advantage of this transformation is that it allows the user to easily see the percentage contribution that each attribute level makes to the improvement in the birth experience, over the bottom state [37].

Results

Study Sample

In total, in the women sample, 1110 subjects completed the questionnaire, 34 in the first phase of the study and 1076 in the second phase of the study (total completion rate 91%), of whom a total of 1097 (100%) passed the rationality test. See Table 2 for details on the demographic and obstetric characteristics of the study population.

Our women sample was very similar to all women who gave birth in The Netherlands in 2012 [38] regarding the mean age (31.6 years vs. 32.1 years), the percentage of primiparous women (47% vs. 48%), and gestational age (<37 weeks, 5% vs. 7%; 37–41 weeks, 93% vs. 90%). The main differences were that our sample included fewer immigrants (10% vs. 23% in The Netherlands in 2012), more home births (27% vs. 17%), fewer secondary-care births (53% vs. 70%), fewer planned cesarean sections (4% vs. 8%), somewhat lower use of pain medication by epidural in the group of women who had a vaginal birth (12% vs. 16%), and fewer complications in the child's health requiring hospitalization (9% vs. 22%).

In the general population sample, 414 persons completed the questionnaire, 85 in the first phase of the study and 329 in the second phase of the study, of whom 367 (89%) subjects passed the rationality test. Compared with data from Statistics Netherlands Centraal Bureau voor de Statistiek (CBS) this sample corresponds very well with the general population in The Netherlands with respect to sex, age distribution, ethnical background, regions in The Netherlands, and level of education [39].

Relative weights of LADY-X attributes

In the MMNL model, all coefficients had P values of less than 0.001 for both samples, except for the difference between the middle and upper levels for the attribute *health care professionals' responses to needs* in the general population sample (see Table 3). In line with the ordering of attribute levels, the utility attached to a woman's birth experience was significantly higher when the levels were more favorable. Greater value was placed on differences between the bottom and middle attribute levels than on those between the middle and top levels (e.g., for the availability attribute, the difference between *rarely* and *most of the time* has greater value than that between *most of the time* and *at all times*), except for the attribute *duration until first contact with child* in the general population sample.

Table 2 – Participant characteristics for both the women and general adult population samples.

Characteristic	Sample women who recently gave birth (N = 1097)	Sample general population (N = 367)
Sex: female	1097 (100)	191 (52)
Age (y), mean ± SD	31.6 ± 4.17	49.7 ± 16.81
Marital status		
Married/cohabiting with partner	1068 (97.4)	248 (67.6)
In a relationship (not cohabiting)	8 (0.7)	20 (5.4)
Single	16 (1.5)	45 (12.3)
Divorced	3 (0.3)	21 (5.7)
Widowed	0 (0)	32 (8.7)
Other	2 (0.2)	1 (0.3)
Ethnical background		
Dutch	993 (90.5)	289 (78.7)
Second-generation immigrants	61 (5.6)	45 (12.3)
First-generation immigrants	43 (3.9)	33 (9.0)
Highest level of education		
Primary school	0 (0)	10 (2.7)
Secondary school	80 (7.3)	137 (37.3)
Vocational education	276 (25.2)	126 (34.3)
Higher level education	436 (39.7)	69 (18.8)
University	305 (27.8)	25 (6.8)
Employment status		
(Self-)employed	922 (84.0)	187 (51.0)
Housekeeping	100 (9.1)	29 (7.9)
Student	9 (0.8)	18 (4.9)
Volunteer work	1 (0.1)	11 (3.0)
Long-term sick leave	14 (1.3)	34 (9.3)
Unemployed	49 (4.5)	24 (6.5)
Retired	0 (0)	63 (17.2)
Other	2 (0.2)	1 (0.3)
Having children	– (–)	275 (74.9)
Number of children, mean ± SD	– ± –	1.38 ± 1.12
Age of youngest child, mean ± SD	– ± –	22.4 ± 15.86
Obstetrical history and characteristics		
History of terminated or extrauterine pregnancy, or miscarriage	307 (28.0)	
Parity: primiparous	499 (46.5)	
Gestational age of newborn (wk), mean ± SD	39.7 ± 1.78	
<22	1 ± 0.1	
22–31	6 ± 0.6	
32–36	48 ± 4.4	
37–41	1008 ± 93.3	
≥42	17 ± 1.6	
Type of pregnancy: singleton	1084 (98.8)	
Place of birth		
Home, led by midwife from primary-care practice	297 (27.1)	
Birth-center, led by midwife from primary-care practice	38 (3.5)	
Hospital, led by midwife from primary-care practice	145 (13.2)	
Hospital, led by secondary-care midwife	214 (19.5)	
Hospital, led by obstetrician	380 (34.6)	
Other	23 (2.1)	
Mode of birth		
Vaginal	846 (77.1)	
Instrumental vaginal (e.g., ventouse or forceps)	115 (10.5)	
Planned cesarean section	40 (3.6)	
Emergency cesarean section	96 (8.8)	
Use of pain medication during labor		
No	714 (65.1)	
Yes, epidural	232 (21.1)	
Yes, other than epidural	135 (12.3)	
Yes, general anesthesia	16 (1.5)	
Transfer during labor	153 (13.9)	
Labor complications concerning the child with hospitalization	100 (9.1)	
Labor complications concerning mother with hospitalization	63 (5.8)	
LADY-X sum score (ranging from 0 to 14), median (range)	12 (0–14)	

Note. All values are n (%) except otherwise indicated.
LADY-X, Labor and Delivery Index.

Preference weights for LADY-X

Separately for each sample, preference weights for LADY-X were derived to represent a utility score between 0 (worst LADY-X scenario) and 1 (best LADY-X scenario) (see Table 4). Utility scores for LADY-X profiles are calculated by summing the preference weight score per level for each attribute.

In the women sample, one attribute clearly resulted in the largest increase in utility (feeling of safety), with a preference weight of 0.188 for the upper level. Five attributes had preference weights of between 0.122 and 0.159 for the upper levels (availability, response to needs, emotional support, concerns, and duration) and one attribute resulted in the lowest increase in utility (information provided) with a preference weight of 0.106 for the upper level. In the general population sample, two attributes resulted in the largest increase in utility (availability and feelings of safety), and four attributes had a preference weight of between 0.120 and 0.129 for the upper level. One attribute clearly results in the lowest increase in utility (duration), with a preference weight of 0.102 for the best level.

The ratio of the upper level weights of both samples provides insights into differences between the women sample and the

Table 3 – Coefficients of panel mixed multinomial logit (MMNL) model by sample.

Attribute		Sample women who recently gave birth (N = 1097)			Sample general population (N = 367)		
		Coefficient	SE	P	Coefficient	SE	P
1. Availability of professionals							
At all times	Mean	0.26	0.05	<0.001	0.48	0.07	<0.001
	SD	0.03	0.08	0.750	0.32	0.09	<0.001
Rarely	Mean	–1.45	0.07	<0.001	–1.09	0.09	<0.001
	SD	0.84	0.05	<0.001	0.80	0.08	<0.001
2. Information provided							
Very well informed	Mean	0.31	0.04	<0.001	0.27	0.06	<0.001
	SD	0.08	0.21	0.705	0.26	0.11	0.019
Inadequately informed	Mean	–0.82	0.04	<0.001	–0.70	0.07	<0.001
	SD	0.07	0.13	0.575	0.37	0.09	<0.001
3. Professionals' response to needs							
Very well responded to	Mean	0.25	0.05	<0.001	0.12	0.07	0.056
	SD	0.03	0.22	0.875	0.38	0.09	<0.001
Not at all responded to	Mean	–1.11	0.05	<0.001	–0.77	0.07	<0.001
	SD	0.44	0.06	<0.001	0.29	0.15	0.049
4. Emotional support by professionals							
Very well supported	Mean	0.22	0.04	<0.001	0.20	0.06	0.001
	SD	0.33	0.07	<0.001	0.02	0.10	0.84
Inadequately supported	Mean	–1.29	0.05	<0.001	–0.83	0.07	<0.001
	SD	0.46	0.06	<0.001	0.57	0.08	<0.001
5. Feelings of safety							
Very safe	Mean	0.28	0.06	<0.001	0.35	0.08	<0.001
	SD	0.21	0.15	0.151	0.42	0.12	0.001
Not safe enough	Mean	–1.8	0.08	<0.001	–1.49	0.10	<0.001
	SD	0.93	0.05	<0.001	0.70	0.08	<0.001
6. Concerns about child's condition							
No concerns	Mean	0.52	0.06	<0.001	0.26	0.08	0.001
	SD	0.12	0.12	0.329	0.57	0.09	<0.001
Many concerns	Mean	–1.22	0.07	<0.001	–0.64	0.08	<0.001
	SD	0.93	0.05	<0.001	0.68	0.08	<0.001
7. Duration until first contact with child							
Not long	Mean	0.71	0.04	<0.001	0.52	0.06	<0.001
	SD	0.53	0.05	<0.001	0.48	0.08	<0.001
Very long	Mean	–0.88	0.06	<0.001	–0.35	0.07	<0.001
	SD	0.49	0.06	<0.001	0.27	0.16	0.083
Model fits							
Log-likelihood function				–8007.6			–3202.5
Akaike information criterion				0.81			0.98
Bayesian information criterion				0.83			1.01
Pseudo R^2				0.41			0.30

general population sample (Table 4). Overall, the ratios are not high, with a maximum of 1.31. The ratio is the highest for the attribute "availability," indicating that the general population assigns greater importance to this attribute than does the women sample. Attributes six and seven have ratios of 0.77 and 0.71, respectively, indicating that the women sample places a higher value on these attributes. Attributes three and four had the lowest ratios, 0.98 and 0.95, respectively, indicating the least difference between the two samples.

Discussion

Main Findings

In the present study, preference weights for the LADY-X classification system were calculated on the basis of two samples, women who had recently given birth and the general population. In line with the ordering of attribute levels, the utility attached to a woman's labor and delivery experience was significantly higher when the levels were more favorable. Greater value was assigned to differences between the bottom and middle attribute levels than to differences between the middle and top levels. This is not surprising because the phrasing of the middle response category is "adequately," which is interpreted as sufficient and thus positive rather than negative and consequently closer to the upper category than to the lower category. The attributes that resulted in the largest increases in utility were "feelings of safety" in the women sample and "feelings of safety" and "availability of professionals" in the general population sample.

Interpretation

In discussing our results, we will focus on two aspects: first, the feasibility of the DCE and, second, the valuation of LADY-X attributes and differences between the two samples. Concerning the feasibility of DCEs, decision tasks are complex and thus not always feasible for individuals lacking experience in the

Table 4 – LADY-X preference weights per attribute level for both samples.

Attributes	Sample women who recently gave birth, levels			Sample general population, levels			Ratio of both samples*
	Best	Moderate	Worst	Best	Moderate	Worst	
1. Availability of professionals	0.145	0.123	0	0.190	0.129	0	1.31
2. Information provided	0.106	0.076	0	0.122	0.087	0	1.15
3. Professionals' response to needs	0.122	0.096	0	0.120	0.096	0	0.98
4. Emotional support by professionals	0.136	0.108	0	0.129	0.100	0	0.95
5. Feelings of safety	0.188	0.161	0	0.216	0.173	0	1.15
6. Concerns about child's condition	0.159	0.097	0	0.122	0.082	0	0.77
7. Duration until first contact with child	0.144	0.075	0	0.102	0.040	0	0.71
Sum	1	0.736	0	1	0.707	0	

LADY-X, Labor and Delivery Index.

* The ratio of the best-level preference weight from the general population sample divided by the best-level preference weight from the women sample.

scenarios considered [37]. A DCE, however, is considered easier than other available valuation methods, for example, the time trade-off and the visual analogue scale [20]. In our study, subjects from both samples indicated that they found it challenging to value the hypothetical scenarios, but completion and consistency rates were nevertheless high. Together with the pretest results, this shows that the DCE on labor experience was feasible, including for subjects who had never given birth or had done so long ago.

We observed differences in the overall ranking of LADY-X attributes. In the women sample, the attributes that were valued highest were "feelings of safety" and "concerns about the child's health," whereas three of the four attributes regarding the health care professionals' behavior exhibited the lowest values (the attributes "information," "needs," and "emotional support"). Compared with the results obtained from the general population, the attribute "availability of health care professionals" and "information provided" had a lower preference weight. In the general population sample, two attributes that had relatively large utilities were "feelings of safety" and "availability of competent health care professionals"; the utilities for the other five attributes did not differ substantially, except for the attribute "duration until first contact with the child," which had a relatively low weight. This attribute, and "concerns about the child," also had a lower preference weight than in the women sample. Explanations for these findings have been provided in the literature. According to Ubel et al. [27], valuation differences between groups can be based on differences in the interpretation of the scenarios or differences in opinions concerning the object of investigation, in this case the birth experience. In line with Bijlenga et al. [20], who also performed valuation studies in multiple samples concerning perinatal care, we expect the differences to stem from differences in interpretation rather than from differences in opinion. Women who have given birth will interpret the scenarios in light of their personal experience, which is not possible, or is possible only to a lesser extent for the members of the general population. For instance, the focusing illusion might have influenced the valuation in the general population [27] such that positive aspects of birth, such as being rewarded afterward by having a child, might have been less apparent for them than for women who had recently experienced a birth. This might have led to the differences between the two samples. The two attributes valued highest at face value ("availability" and "feelings of safety") may appear the most important to someone who has not recently given birth, whereas "concerns about the child," as well as "duration until the first contact," may (in retrospect) not be deemed particularly important, particularly when assuming that the child will be healthy afterward, in contrast to the view of a woman who recently experienced giving birth.

Strength and Limitations

One strength of this study is the design we used to estimate preference weights, which followed state-of-the-art standards and exploited a pilot study to optimize the design of the main study [29]. Another strength is that results were obtained for two samples. Moreover, the DCE was carefully pretested to improve its feasibility and to make it applicable for both women who had experienced birth themselves and for members of the general population who did not (recently) have this experience.

The main limitation of our study is that the women sample was not representative of the population of women who give birth in The Netherlands in all respects, as the comparison with data from the national population showed. It appears that our recruitment strategy attracted more women with a home birth and fewer women who themselves or whose children had experienced complications during or after the birth, or first- and second-generation immigrants. It would be worthwhile to investigate whether references for these specific subgroups differ.

Further Research

Studying possible interaction effects of the seven LADY-X attributes was not the focus of this study and was not feasible using our data but might be relevant to gain a complete understanding of the mutual relationship among the seven LADY-X attributes and thus should be the focus of future research. In addition, studying preference heterogeneity was beyond the scope of this article, but it might be relevant to assess whether differences in valuation based on demographic or obstetrical characteristics and personal birth experiences exist.

It should be stressed that because of the focus on intrapartum care, LADY-X is not expected to be sensitive to antepartum care, to the postnatal phase or to neonatal outcomes, because no health outcomes for the mother or for the child are included. Therefore, the development of utility measures for antenatal and postnatal care aspects is still recommended as a subject for future research.

Implications for the Use of LADY-X

Decisions in the intrapartum period are preference-sensitive. However, group-level results on cost-effectiveness are relevant for decisions in this period at a policy or service level, for

example, about the method of pain relief (see, e.g., the study of Freeman et al. [40]). If the cost-effective intervention is implemented in practice, individual heterogeneity can be addressed if the costs of the implemented intervention are comparable to or higher than the costs of the intervention that is preferred by the woman. Group-level results may indicate that most of the women prefer the more expensive treatment in this case, but a woman is allowed to choose differently on the basis of her own preferences. This is comparable to policy decisions using other utility measures, for example, the EuroQol five-dimensional questionnaire (see, e.g., Van den Hout et al. [41]). If the cost-effective intervention is less expensive than the intervention that is preferred by the individual woman, however, there is a fundamental conflict between preference-sensitivity and cost-effectiveness. In this case, no general conclusion can be drawn; the decision will be dependent on the context, for example, the reimbursement system.

In the present study, we assessed preferences in two samples, of patients and of the public, because no conclusive answer has been given to the question of whose preferences to use [21–27]. For economic evaluations from a societal perspective, the values of a well-informed public should be used because it is the public's resources that are allocated. Furthermore, it is argued that the public provides rational valuations because it does not have a direct interest in the treatment of a particular health state. Patients' preferences, however, might be used because it has proven difficult to obtain a well-informed public [26,27]. Patients are more knowledgeable of their situation than individuals who attempt to imagine the situation [23]. In conclusion, in line with Gold et al. [25], we recommend the use of preference weights obtained from the women sample for the calculation of standard utility scores for economic evaluations that compare multiple existing intrapartum care interventions or new interventions with existing interventions within the perinatal setting. In contrast, for resource allocation purposes in a broader context, the public preferences should be used.

Given the results of this study, we regard LADY-X as a suitable utility measure for use in economic evaluations of perinatal care interventions. It should be mentioned that with the preference weights available, LADY-X can be used as a condition-specific utility measure that represents the subjective evaluation of the key aspects of labor and delivery. If differences in the health outcomes of the mother and/or child are expected in an economic evaluation, generic quality-of-life measures should be used because LADY-X is restricted to the labor and birth experience and generic outcome measures are preferred because of their broad comparability. In this case, LADY-X might be a suitable secondary outcome measure.

Conclusions

We believe that LADY-X satisfies the need for a utility instrument for intrapartum care. The LADY-X questionnaire, now also available with preference weights, offers the opportunity to assess women's birth experiences and include utility scores for these experiences in economic evaluations when generic outcome measures are not applicable.

Source of financial support: This study was supported by a grant from The Netherlands Organization for Health Research and Development (Nederlandse organisatie voor gezondheidsonderzoek en zorginnovatie, ZonMw) (project no. 171102017).

Supplemental Materials

Supplemental materials accompanying this article can be found in the online version as hyperlink at http://dx.doi.org/10.1016/j.jval.2015.07.005 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Schroeder E, Petrou S, Patel N, et al. Cost effectiveness of alternative planned places of birth in woman at low risk of complications: evidence from the Birthplace in England national prospective cohort study. BMJ 2012;344:e2292.
[2] Palencia R, Gafni A, Hannah ME, et al. TermBreech Trial Collaborative Group. The costs of planned cesarean versus planned vaginal birth in the Term Breech Trial. CMAJ 2006;174:1109–13.
[3] Borup L, Wurlitzer W, Hedegaard M, et al. Acupuncture as pain relief during delivery: a randomized controlled trial. Birth 2009;36:5–12.
[4] Freeman LM, Bloemenkamp KW, Franssen MT, et al. Remifentanil patient controlled analgesia versus epidural analgesia in labour: a multicentre randomized controlled trial. BMC Pregnancy Childbirth 2012;12:63.
[5] Bakker JJ, De Vos R, Pel M, et al. Start of induction of labour with oxytocin in the morning or in the evening: a randomised controlled trial. BJOG 2009;116:562–8.
[6] Kershaw K, Jolly J, Bhabra K, et al. Randomised controlled trial of community debriefing following operative delivery. BJOG 2005;112:1504–9.
[7] Sluijs AM, Cleiren MP, Scherjon SA, et al. No relationship between fear of childbirth and pregnancy-/delivery-outcome in a low-risk Dutch pregnancy cohort delivering at home or in hospital. J Psychosom Obstet Gynaecol 2012;33:99–105.
[8] Petrou S, Henderson J. Preference-based approaches to measuring the benefits of perinatal care. Birth 2003;30:217–26.
[9] Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
[10] Gartner FR, Freeman LM, Rijnders ME, et al. A comprehensive representation of the birth-experience: identification and prioritization of birth-specific domains based on a mixed-method design. BMC Pregnancy Childbirth 2014;14:147.
[11] Lancaster K. A new approach to consumer theory. J Pol Econ 1966;74:134–57.
[12] Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy 2003;2:55–64.
[13] Ryan M, Netten A, Skatun D, et al. Using discrete choice experiments to estimate a preference-based measure of outcome—an application to social care for older people. J Health Econ 2006;25:927–44.
[14] Clark MD, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics 2014;32:883–902.
[15] de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ 2012;21:145–72.
[16] Ratcliffe J, Longworth L. Investigating the structural reliability of a discrete choice experiment within health technology assessment. Int J Technol Assess Health Care 2002;18:139–44.
[17] Hundley V, Ryan M, Graham W. Assessing women's preferences for intrapartum care. Birth 2001;28:254–63.
[18] Hundley V, Ryan M. Are women's expectations and preferences for intrapartum care affected by the model of care on offer? BJOG 2004;111:550–60.
[19] Bijlenga D, Birnie E, Mol BW, et al. When outcome is a balance: methods to measure combined utility for the choice between induction of labour and expectant management in mild risk pregnancy at term. BMC Pregnancy Childbirth 2007;7:10.
[20] Bijlenga D, Birnie E, Mol BW, et al. Obstetrical outcome valuations by patients, professionals, and laypersons: differences within and between groups using three valuation methods. BMC Pregnancy Childbirth 2011;11:93.
[21] Boyd NF, Sutherland HJ, Heasman KZ, et al. Whose utilities for decision analysis? Med Decis Making 1990;10:58–67.
[22] Brazier J, Akehurst R, Brennan A, et al. Should patients have a greater role in valuing health states? Appl Health Econ Health Policy 2005;4:201–8.
[23] Brazier J. Valuing health states for use in cost-effectiveness analysis. Pharmacoeconomics 2008;26:769–79.
[24] Dolan P. Whose preferences count? Med Decis Making 1999;19:482–6.
[25] Gold MR, Patrick DL, Torrance GW, et al. Identifying and valuing outcomes. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. Cost-effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
[26] Stiggelbout AM, deVogel-Voogt E. Health state utilities: a framework for studying the gap between the imagined and the real. Value Health 2008;11:76–87.
[27] Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Qual Life Res 2003;12:599–607.

[28] Hensher DS, Rose JM, Greene WH. Applied Choice Analysis: A Primer. Cambridge, UK: Cambridge University Press, 2005.
[29] Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16:3–13.
[30] Hall J, Fiebig DG, King MT, et al. What influences participation in genetic carrier testing? Results from a discrete choice experiment. J Health Econ 2006;25:520–37.
[31] Bliemer MCJ, Rose JM. Experimental design influences on stated choice outputs: an empirical study in air travel choice. Transport Res Part A Policy Pract 2011;45:63–79.
[32] Carlsson F, Martinsson P. Design techniques for stated preference methods in health economics. Health Economics 2003;12:281–94.
[33] Rose JM, Bliemer MCJ. Constructing efficient stated choice experimental designs. Transp Rev 2009;29:587–617.
[34] Willis GB. Cognitive interviewing: a tool for improving questionnaire design. In: Cognitive Interviewing in Practice: Think-Aloud, Verbal Probing and Other Techniques. Thousand Oaks, CA: Sage, 2005.
[35] Hoefman RJ, van Exel J, Rose JM, et al. A discrete choice experiment to obtain a tariff for valuing informal care situations measured with the CarerQol instrument. Med Decis Making 2014;34:84–96.
[36] Coast J, Flynn TN, Natarajan L, et al. Valuing the ICECAP capability index for older people. Soc Sci Med 2008;67:874–82.
[37] Al-Janabi H, Flynn TN, Coast J. Estimation of a preference-based carer experience scale. Med Decis Making 2011;31:458–68.
[38] Stichting Perinatale Registratie Nederland. Perinatale Zorg in Nederland 2010. Utrecht, The Netherlands: Stichting Perinatale Registratie Nederland, Available from: http://statline.cbs.nl/Statweb/. [Accessed: December, 2013].
[39] CBS (Statistics Netherlands). Statline. Available from: http://statline.cbs.nl/Statweb/. [Accessed: December, 2013].
[40] Freeman LM, Bloemenkamp KW, Franssen MT, et al. Patient controlled analgesia with remifentanil versus epidural analgesia in labour: randomised multicentre equivalence trial. BMJ 2015;350:h846.
[41] van den Hout WB, Peul WC, Koes BW, et al. Prolonged conservative care versus early surgery in patients with sciatica from lumbar disc herniation: cost utility analysis alongside a randomised controlled trial. BMJ 2008;336:1351–4.